

Overview of quantitative news interpretation methods applied in financial market predictions

Miklós Vázsonyi

Received 2010-01-05

Abstract

This paper describes currently known methods of quantitative news interpretation applied in financial market predictions. Brief summaries are made regarding all the listed methods of automatic news interpretation, some commercial applications are mentioned and finally a conclusion is drawn about the usability and prospects of quantitative news analysis with statistical machine learning methods. The aim of this paper is to provide an overview on the related research activities performed so far and explore further research directions to improve the predictive capability of currently known methods.

Keywords

quantitative news interpretation · statistical machine learning · financial market prediction.

Acknowledgement

The author is grateful to Prof. Csaba Pléh and Dr. Ferenc Kiss from Budapest University of Technology and Economics for their professional support. The research activity was funded by the Hungarian Eötvös Scholarship hosted by the Hungarian Scholarship Board. The research was conducted at the University of Cyprus – HERMES European Center of Excellence on Computational Finance and Economics under the professional supervision of Prof. Hercules Vladimirov.

Miklós Vázsonyi

PhD School of Economics and Organizational Sciences, BME, H-1521 Budapest, Műegyetem rkp. 9., Hungary
e-mail: miklos@vazsonyi.com

1 Motivation

Financial market prediction methods generally use quantitative values to provide point or interval estimations on returns, volatility or trade volume. Historized timeseries of financial products with previous market prices are commonly used in portfolio optimization. However, quantitative analysis of qualitative economic news given in unstructured textual formats started a couple of years ago when Internet news media became the dominant information source of investors. Recent studies identify systematic relationships between trading volume and measures of communication activity [1].

If an investor has some private information about the value of an asset, his trades reveal information to the market. In equilibrium, the sensitivity of prices to trades will depend on the prevailing level of information asymmetry. This concept was formalized and since that time many researchers have studied the link between information asymmetry and the impact of trading on prices. Since investors differ in their abilities to interpret the news, the release of public information may actually increase information asymmetry of market participants [7].

If we empirically analyse the significant movements of firms' stock prices then we realize that they do not seem to correspond to changes in quantitative measures of firms' fundamentals. Consequently, qualitative analysis may help explain stock returns. Analyzing a more complete set of events that affect firms' fundamental values can lead to identify common patterns in firm responses and market reactions to events. Textual news is a potentially important source of information about firms' fundamental values. Very few stock market investors directly observe firms' production activities, they get most of their information secondhand. Their three main sources are analysts' forecasts, quantifiable publicly disclosed accounting variables, and linguistic descriptions of firms' current and future profit-generating activities. Future research on quantifying textual news has the potential to improve our understanding of how information is incorporated in asset prices [2].

To predict the impact of macroeconomic news on stock prices is not an easy task. News about faster growing economy usually creates expectations of higher corporate earnings and div-

dividends. These expectations in turn should boost stock prices, since stock's price should match the expected stream of future dividends from that stock, discounted to their present value. However, news that the economy is growing faster than anticipated will also lead to higher expected interest rates – the rates used to discount future dividends. Whether stock prices in fact rise or fall will depend on whether the stream of expected future dividends or the discount rate plus compensation for risk responds more strongly to the news. The same rule applies to the response of stock prices to news of inflation: inflation boosts prospective nominal future earnings and the nominal rate at which such earnings are discounted. Given the uncertain interplay of these variables, it is not surprising that many studies cannot identify consistent effects of macroeconomic news on stock prices [3].

If market participants disagree about the effects of surprises in announcements there should be increased trading activity in the market soon after the announcements. In contrast, if they are in consensus about the effects of new information, trading activity may not be abnormal even when prices change. Thus, examining the trading activity provides useful information about the actions taken by the market participants based on news that stock returns alone cannot [8].

If results showed that financial news announcement have a measurable impact on financial asset prices it would contradict the so-called Efficient Markets Hypothesis that states that asset prices fully reflect all available information, however, it would support the idea of cognitive investor sentiment and over- and under-reaction to news content.

2 Methods of quantitative news interpretation in financial market predictions

Currently known methods of quantitative news interpretation are described below with reference to the original papers. Methods were selected according to their relevance and prior focus to financial optimization efforts. Description of other methods of quantitative content analysis or non-news based sentiment analysis are not in the scope of the current paper.

2.1 News cardinality analysis

Early studies about news impact measurement on financial market calculated only with the varying cardinality of the news available for investors. Most of the research show that there is a positive correlation between news cardinality and trade volume plus volatility.

X. Liang defined the web stock news volumes (WSNV) indicator and measured the impact of it on financial market behavior (see paper [11] from 2006). He downloaded news items from www.yahoo.com, quote.com and www.smartmoney.com on a daily basis with internet crawler engines. He classified news according to direct and indirect aspects: direct news comes from a dedicated site containing the current news of a certain company (e.g. <http://finance.yahoo.com/q/h?s=amzn>); indirect

news: company is mentioned on an other news site. His main conclusion is that significant increases of web stock news volumes are linked with the significant changes of stock prices [11].

2.2 Pessimism factor analysis

Paul C. Tetlock showed measurable correlation between written financial media content and aggregate financial market performance as an evidence that news content can predict movements in stock market activity (see paper [1] from 2007). This method focuses on the measurement of immediate influence of the Wall Street Journal's (WSJ) „Abreast of the Market” column on the daily US stock market returns.

Principal component analysis is used to create a simple measure of media pessimism from WSJ news of 16 years then the intertemporal links between this measure of media pessimism and the stock market is estimated using basic vector autoregressions (VARs). 77 predefined word categories of Harvard psychosocial dictionary is used to find the category with the highest variance which is called the pessimism factor. This method follows a bag-of-words approach to handle textual documents. Conclusions of the pessimism factor research are:

- High levels of media pessimism robustly predict downward pressure on market prices on the next trading day, followed by a reversion to fundamentals within one week.
- Unusually high or low values of media pessimism forecast high market trading volume.
- Low market returns lead to high media pessimism.
- The changes in market returns that follow pessimistic media content are dispersed throughout the trading day, rather than concentrated after the release of information.

Tetlock defines two theories based on extremes content of news to justify the results described above:

- The sentiment theory predicts that short-horizon returns will be reversed in the long run (content is pure noise).
- The information theory predicts that short-horizon returns will persist indefinitely (content is pure information).

Tetlock proposes that a computer program could calculate the daily value of pessimism and use predetermined coefficients, derived from predictability regressions to forecast future returns. Depending on whether this forecast is positive or negative, the media-based trading strategy would go long or short on the Dow Jones index. To check his assumptions, Tetlock used a time series of daily returns from 1 January 1984 to 17 September 1999 from the Wharton Research Data Services' access to the historical Dow Jones Industrial Averages. Experimental research shows that the pessimism media factor exerts a statistically and economically significant negative influence on the next day's returns. Also, analysis of data showed that this negative influence is only temporary and is almost fully reversed later in the trading

week. The evidence of an initial decline and subsequent reversal is consistent with neither the new information nor the stale information theories of the newspaper column. If the column contained new information about fundamentals, there could be an initial decline in returns, but this would not be followed by a complete return reversal. If the column contains only information already incorporated into prices, media pessimism would not significantly influence returns. The evidence is consistent with temporary downward price pressure caused by pessimistic investor sentiment [1].

Most of the financial market theories say that media pessimism should have no effect on future market activity because expectations are already incorporated into prices. On the other hand, if we believe that the media pessimism measure contains no information about past, present, and future, then one would not expect to observe any impact of pessimism on market activity [1].

2.3 Method of relative frequencies of negative words

Paul C. Tetlock, Maytal Saar-Tsechansky and Sofus Macskassy described a news-based automated trading strategy based on relative occurrence of negative words in firm specific financial news in an effort to predict firms' accounting earnings and stock returns (see paper [2] from 2008). The research used news stories about S&P 500 firms from 1980 through 2004 provided by Dow Jones News Service (DJNS). A simplified bag-of-words representation was used to interpret textual data according to the relative frequency of negative words defined by the Harvard psychosocial dictionary. Three main findings of their research effort are:

- The fraction of negative words in firm-specific news forecasts low firm earnings, stock market prices respond to the information embedded in negative words with a small, one-day delay.
- Firms' stock prices briefly underreact to the information embedded in negative words.
- The earnings and return predictability from negative words is largest for the stories that focus on fundamentals.

Together these findings suggest that linguistic media content captures otherwise hard-to-quantify aspects of firms' fundamentals, which investors quickly incorporate into stock prices. Potential profits are plausible from using daily trading strategies based on the relative frequency of negative words in a continuous intraday news source. Negative words in stories about fundamentals predict earnings and returns more effectively than negative words in other stories. The following data sources were used during the research:

- Center for Research on Security Prices (CRSP): S&P index constituents and their stock price data;
- CRSP company name change file: to identify situations in which a firm changed its name;

- Institutional Brokers' Estimate System (I/B/E/S): analyst forecast information;
- Compustat: accounting information;
- Factiva database: news stories.

Tetlock and his colleagues implemented an automated story retrieval system. For each S&P 500 firm, the system constructed a query that specifies the characteristics of the stories to be retrieved. The system then submitted the query and recorded the retrieved stories. In total, they retrieved over 350,000 qualifying news stories. Each of the stories met certain requirements that eliminated irrelevant stories. Their study shows that negative words have better predictive power than any other single category, including positive words, in other terms negative words have a much stronger correlation with stock returns than other words. These results are also consistent with a large body of literature in psychology which argues that negative information has more impact and is more thoroughly processed than positive information across a wide range of contexts. They also showed that news stories concentrate around earnings announcement days. This finding suggests that news stories could play an important role in communicating and disseminating information about firms' fundamentals [2].

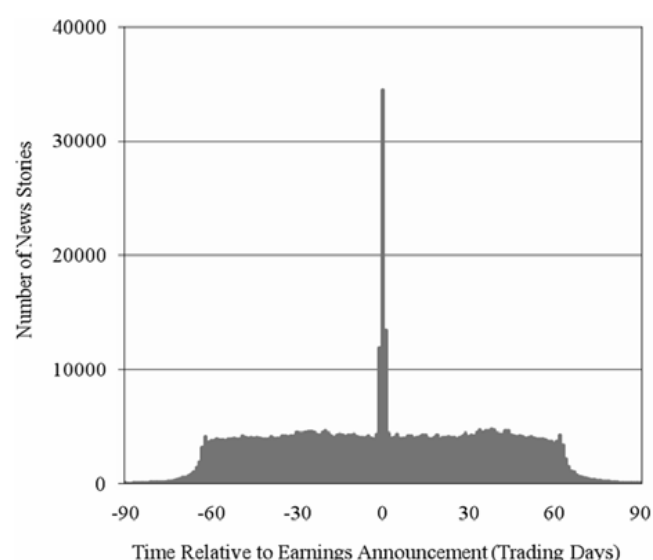


Fig. 1. Number of published news as a function of days around earnings announcement date [2]

Before counting instances of negative words, all qualifying news stories were combined for each firm on a given trading day into a single composite story. The fraction of negative words was standardized in each composite news story by subtracting the prior year's mean and dividing by the prior year's standard deviation of the fraction of negative words. Formally, two measures of negative words were used [2]:

$$Neg = \frac{\text{No. of negative words}}{\text{No. of total words}}$$

$$neg = \frac{Neg - \mu_{Neg}}{\sigma_{Neg}}$$

where μ_{Neg} is the mean of Neg and σ_{Neg} is the standard deviation of Neg over the prior calendar year. The standardization may be necessary if Neg is nonstationary, which could happen if there are regime changes in the distribution of words in news stories – for example, the DJNS or WSJ changes its coverage or style. Based on these relative frequencies of negative words an automated news-based trading strategy was proposed as follows:

- All firms with positive DJNS news stories from 12:00 am to 3:30 pm on the prior trading day were classified into the long portfolio;
- All firms with negative stories were classified into the short portfolio;
- Both the long and short portfolios were hold for 1 full trading day and rebalance at the end of the next trading day.

Ignoring trading costs, the cumulative raw returns of this long-short strategy would be 21.1% per year. This ideal case is distorted when we calculate with different transactional costs. The estimated impact of reasonable transaction costs on the trading strategy's profitability is shown in the table below [2].

Tab. 1. Effect of transactional costs on the returns of the proposed trading strategy [2]

Trading Costs (bps) (round-trip trade)	Raw Annualized Returns (%)
0	21.07
1	18.25
2	15.49
3	12.80
4	10.17
5	7.60
6	5.09
7	2.64
8	0.25
9	2.09
10	4.37

From the analysis above, it turns out that negative words in firm-specific stories leading up to earnings announcements significantly contribute to a useful measure of firms' fundamentals. The main result is that negative words in firm-specific news robustly predict slightly lower returns on the following day [2].

2.4 Surprise analysis of publicly known versus newly announced macroeconomic and political data

Zeynep Önder and Can Simga-Mugan analysed the impact of political and macroeconomic news in emerging markets: Argentine (Buenos Aires Stock Exchange (BASE)) and Turkey (Istanbul Stock Exchange (ISE)) to investigate the origin of high returns (see paper [4] from 2006). Higher uncertainty relative to developed markets increases both risk and return. Two main sources of uncertainty are politics and economics. They examined the effects of macroeconomic and political news items on the volatility of returns and total trading volume between 1995

and 1997 (Economic and political news items were downloaded from Wall Street Journal and New York Times databases.). Their main conclusions were (some further details can be found in Appendix 1) [4]:

- Both economic and political factors, as well as specific market characteristics, should be taken into consideration by investors when making investment decisions in emerging markets.
- Political news and world economic news increase volatility in both markets.
- Political news decrease trading volume in the BASE but increase it in the ISE.
- There is a positive and significant correlation between world economic news items and volume in Argentina, and a positive association between domestic and world economic news and volume in the Turkish market.

Prior studies also investigated the effect of several economic announcements on stock returns (Jain 1988; Mitchell and Mulherin 1994; Pearce and Roley 1985) as well as interest rate and foreign exchange markets (Ederington and Lee 1993; Tanner 1994). Harvey (1995) comprehensively analyses twenty emerging markets, for which he forecasts returns using both world and local economic information. The results show that local information strongly influences returns in these markets [4].

Several studies have recognized that political information affects the stock market (e.g., Gartner and Wellershoff 1995; Hensel and Ziemba 1995; Herbst and Slinkman 1984; Huang 1995; Lobo 1999; Riley and Luksetich 1980). Most of these studies examine the effect of presidential and midterm elections, and the result of elections, on returns in U.S. markets, finding noticeable relations [4].

Cutler et al. (1989) first relate the stock returns to macroeconomic indicators, then examine whether the remaining return variation can be explained by "identifiable world news" reported in the business section of the New York Times from 1941 to 1987. The authors find the effect of such news to be "surprisingly small" [4].

A couple of years later Leonardo Bartolini, Linda Goldberg and Adam Sacarny analysed the effects of news of macroeconomic data on asset price changes (see paper [3] from 2008). Governments and some private organizations regularly issue statistics on the performance of the nation's economy. The nature and extent of the market response will vary with the news announcement. By "news," they mean the surprise element, or the difference between the actual value announced for an indicator and market participants' prior expectation of what that value would be. (The expected value is captured by the median response from the last preceding weekly survey of market participants conducted by Bloomberg L.P.) The main conclusions of their research are (some further details can be found in Appendix 2) [3]:

- Only a few announcements – the nonfarm payroll numbers, the GDP advance release, and a private sector manufacturing report – generate price responses that are economically significant and measurably persistent. Bond yields and the exchange value of the US dollar show the strongest response and stock prices the weakest.
- The strongest effects are seen on interest-bearing assets, and the weakest and most erratic on stock prices: unexpected changes in the data generally have the most marked impact on interest rates, a weaker impact on exchange rates, and an even weaker impact on equity prices.
- Indicators such as the government statistics on personal income and personal consumption expenditures excluding food and energy typically have a small and transitory impact on prices.
- The significant responses support the view that asset prices rise (in nominal terms) in response to news of stronger growth and faster inflation.
- while the direction and size of news effects on asset prices tend to be consistent from the time of the release to the end of the day, the immediate impact can generally be measured more precisely than the full-day impact because the accumulation of other shocks to asset prices through the business day makes the identification of persistent effects more difficult.

T. Clifton Green studied the impact of macroeconomic news on government bond prices (see paper [7] from 2004). They concluded the following statements (some further details can be found in Appendix 3):

- The results show a significant increase in the informational role of trading after economic news stories announcements.
- The informational role of trading is greater after announcements with a larger initial price impact, and the relation is associated with the surprise component of the announcement and the precision of the public information.
- The results provide evidence that government bond order flow reveals fundamental information about riskless rates.

Prem C. Jain investigated the effects of money supply, consumer price index (CPI), producer price index, industrial production, and the unemployment rate related news announcements on stock prices from 1978 to 1984 (see paper [8] from 1988). The empirical results indicated the following:

- Surprises in announcements about money supply and CPI are significantly associated with stock price changes (1-percentage-point surprise in the CPI results in a decline in stock prices of about 0.55%). The announcements of the other three variables do not affect stock prices significantly.

- Trading volume is not affected by any of the five economic variable announcements, indicating that market participants do not differ substantially in the interpretations of the effects of announcements.
- The speed of adjustment analysis indicates that the effect of information on stock prices is reflected in a short period of 1 hour.

Pierluigi Balduzzi, Edwin J. Elton, and T. Clifton Green measured the intraday effects of macroeconomic news on bond prices and trade volume between July 1, 1991 and September 29, 1995 (see paper [12] from 2001) (some further details can be found in Appendix 4). They conclude that:

- public macroeconomic news can explain a substantial fraction of price volatility in the aftermath of announcements,
- the adjustment to news generally occurs within one minute after the announcement so the public news tends to be incorporated very quickly into prices.

2.5 Unsupervised clusterization of firm specific news as good or bad based on market activity

Moshe Koppel and Itai Shtrimberg realized that models based on lexical features can distinguish good news from bad news with accuracy of about 70% (see paper [5] from 2006). A simple and novel method for generating labeled examples for sentiment analysis was introduced: news stories about publicly traded companies are labeled positive or negative according to price changes of the company stock. It is shown that there are many lexical markers for bad news but none for good news. Overall, learned models based on lexical features can distinguish good news from bad news with accuracy of about 70%. Unfortunately, this result does not yield profits since it works only when stories are labeled according to contemporaneous price changes but does not work when they are labeled according to subsequent price changes. The main conclusions of their research are the following [5]:

- There are a number of features that are clear markers of negative documents. These include words such as shortfall, negative and investigation. Documents in which any of these words appear are almost always negative. The first twenty words with highest information gain all are negative markers.
- There are no markers of positive stories.

The novel idea here is the automatic unsupervised clusterization of large amount of news. The use of price movements correlated with the appearance of news items is a promising method for automatically generating a labeled corpus without directly invoking individual human judgments (though, of course, stock movements themselves are a product of collective human judgment). In this work, there are no assumptions by making judgments regarding a story itself. Only the reaction of the market to the story is important. The use of stock price movements offers several advantages over hand-labelled corpora:

- The entire procedure is automatic and thus a large labelled corpus can easily be generated.
- The collective judgment of the market is a more reliable determiner of sentiment than that of a small number of judges. The market effectively serves as judge.

News stories concerning each of the stocks in the S&P500 index were gathered for the years 2000-2002. The stories were taken from the Multex Significant Developments corpus. The total number of stories were over 12,000 – an average of 24 stories per stock. The average length of a story was just over 100 words. Definition of stories according to positive and negative labels are:

- Positive: the stock in question rose 10% or more;
- Negative: the stock declined 7.8% or more.

Rather high percentage thresholds were used because such dramatic price moves can be safely assumed to be reactions to news stories and not mere reflections of general market moves or random fluctuation. The lower threshold for downward moves was chosen so as to provide an equal number of negative examples as positive examples. Each text was represented as a binary vector reflecting whether or not a feature is present in the story, but ignoring frequency. The categorization methodology consisted of selecting the 100 features with highest information gain in the training corpus and then using a linear SVM (and other learners) to learn a model [5].

Rohitha Goonatilake and Susantha Herath focus on the effect of news that surfaces throughout the day in the stock market (see paper [9] from 2007). News stories about publicly traded companies were labeled positive, negative or neutral according to price changes in the company stock. They statistically analyzed the influence of news on the DJIA, NASDAQ, and S&P 500. From the analysis of the data collected over a ten-week period of time they concluded the following statements (some further details can be found in Appendix 5) [9]:

- At the 5% significance level, the data provides sufficient evidence to determine that there is an association between the nature of the number of news items and the market fluctuations, measured by an increase, a decrease, or unchanged.
- On a given day, if the stocks have an upward tendency, then all three indicators will have the same trend, and vice versa to conclude that if the buyers want to buy, they go on a buying spree and DJIA, NASDAQ, S&P 500, and crude oil prices share the same time-series configuration.
- A set of regression models can be obtained that predicts crude oil prices from the DJIA and vice versa with a strong correlation coefficient.
- If a major news story continues even after it has first surfaced, it can still affect the market for a period and diminishes after some time.

Satoru Takahashi, Masakazu Takahashi, Hiroshi Takahashi and Kazuhiko Tsuda used Naive Bayes classifier for text labeling (see paper [10] from 2007). They measured stock price change before and after publishing news. They performed morphological analysis and pattern matching to extract keywords then good, bad and neutral news clusters were created. The text categorization system had about 80% of classification accuracy [10].

X. Liang used web stock news volumes analysis along with good/bad classification of news item (see paper [11] from 2006). He stated the following conclusions [11]:

- In the bear market, bad information influences the market more severely than good information. Namely, the asymmetry phenomenon should be addressed separately for the bear market.
- The underreaction is observed in which if the news is good, the stock prices keep the tendency up following the first optimistic reaction, and if the news is bad, prices keep the tendency down following the first pessimistic reaction
- News with less probability to occur insinuates a larger impact on the stock market. As a matter of fact, the introduction of news entropy has a potential to solve this question.

2.6 Headline analysis

Niederhoffer (1971) explored the effect of headline news appearing in the New York Times and the Los Angeles Times on the stock market from 1950 to 1966. The author groups headline news into various classes, and as good or bad [4].

Wesley S. Chan used a comprehensive database of news headlines about individual companies and examined monthly returns following public news. He compared them to stocks with similar returns but no identifiable public news. He found strong drift after bad news (see paper [6] from 2003). Major conclusions are:

- stocks with bad public news display a negative drift for up to 12 months. Less drift is found for stocks with good news.
- stocks that had no news stories in the event month tend to reverse in the subsequent month. The reversal is statistically significant, even after controlling for size and book-to-market.
- Underreaction to news seems stronger in lower priced, more illiquid stocks.

I carried out a headline news analysis based market impact calculation in the following setup. I collected news items for S&P500 companies between 1 January 1987 and 1 July 2009 from the Associated Press Financial Wire (from the LexisNexis online news database) where the name of the company appears in the headline of the news (90,226 news items). I performed sentence tokenization on the corpus, stopword elimination and basic stemming on unique words. I defined typical words the following four categories:

- Positive nouns and verbs;
- Negative nouns and verbs;
- Positive adjectives;
- Negative adjectives.

I investigated the cardinality of expressions consisting adjective + noun/verb pairs. This way there are four different expression types: PP (positive adjective+positive noun/verb), PN, NN, NP respectively. Having this measure for every headline news items I classified the news as good or bad. Then I investigated the returns on the publishing dates of these headline news items and classified them into positive and negative return days. My analysis resulted in relatively stable ~80% precision and recall regarding the predictive capabilities of the method. Some further details can be found in Appendix 6.

2.7 Natural language processing approaches

Taeho Hong and Ingoo Han developed a competency map based textual data miner application to support financial prediction using 180,000 news stories and preliminary expert knowledge base containing knowledge of domain experts (see paper [13] from 2004).

Prior knowledge base is built by the codification of domain experts and used to support automated news interpretation. News is represented in trigram model extended with competency maps [13].

Robert P. Schumaker and Hsinchun Chen used Bag of Words, Noun Phrases, and Named Entity representation of news to quantify language having 9,211 financial news articles and 10,259,042 stock quotes covering the S&P 500 stocks during a five week period (see paper [14] from 2009). News articles were obtained from Yahoo Finance using companies' stock ticker symbols. In this study Noun Phrasing were accomplished through the use of a syntax where parts of speech are identified through the aid of a lexicon and aggregated using syntactic rules on the surrounding parts of speech, forming noun phrases. In the Named Entities representation the text processing is built upon Noun Phrases by using lexical semantic/syntactic tagging where nouns and noun phrases can be classified under predetermined categories. Named Entity categories include date, location, money, organization, percentage, person and time. Typically, successful taggers have large lexicons of sample entities and/or word patterns, which may include both syntax and lexical information. Theoretically, Named Entities provide the smallest coverage of a text document. Machine learning algorithms (support vector machine) were trained on quantitative training sets. The system is achieving fairly reasonable results at a 2.09% return over a year of experiments. The authors also concluded that industrial sector based training had the best performance [14].

3 Commercial applications

Some commercially available applications are listed below that performs media sentiment analysis through automated fi-

nancial news interpretation. Surely, it is not a complete list of available services, however, it is clear that more companies started to commercialize the idea of quantitative news interpretation in financial market prediction.

3.1 Reuters Sentiment Engine

Reuters NewsScope Sentiment Engine processes a stream of Reuters news items, producing sentiment data for a list of customer determined target companies. Scores consist of how positive, negative, or neutral a particular article, or substantive entity within an article, is written by the author. The engine can score many articles per second making it applicable to real time algorithmic trading as well as a host of other use cases [16].

3.2 MediaSentiment engine

MediaSentiment Pro includes the advanced search engine that generates Buy/Sell signals based on sentiment analysis of earnings releases. MediaSentiment Inc. has developed a technology that can make a rapid assessment of the sentiment expressed by the management in the earnings news release and correlate that with Wall Street analysts' estimates and the overbought/oversold signals in the trading market to uncover profitable trading opportunities [17].

3.3 YellowBrix Sentiment Analytics

YellowBrix Sentiment Analytics provides an aggregated view of media sentiment and news analytics from global newspapers, news wires, trade journals and business blogs, providing information for investors. Visual representation and interactive charting of media sentiment across multiple tickers is part of the service, indices, and customized watch lists are available to quickly assess the media sentiment benchmarked against stock performance. It provides news headlines with at-a-glance summaries. It analyzes industry and business intelligence using measurement and sentiment analytics for the volume of media exposure for companies, people, products and topics in the mainstream news media and blogs; and uses sentiment analysis to weigh the level of positive or negative content [18].

3.4 Dow Jones Economic Sentiment Indicator

The Dow Jones Economic Sentiment Indicator aims to predict the health of the U.S. economy by analyzing the coverage of 15 major daily newspapers in the U.S. It uses a numerical scale from 0 to 100 to express the balance of sentiment in articles about the economy. Its back-testing to 1990 shows that it clearly highlighted the risk that the U.S. economy was sliding into recession in 2001 and 2008 and suggests the indicator can help predict economic turning points as much as seven months in advance of other indicators. The Dow Jones Economic Sentiment Indicator is calculated using a proprietary algorithm through Dow Jones Insight, a media tracking and analysis tool. Dow Jones Insight's global content collection includes

Fig. 2. Architecture of the expert system based Knowledge-Based News Miner system [13]

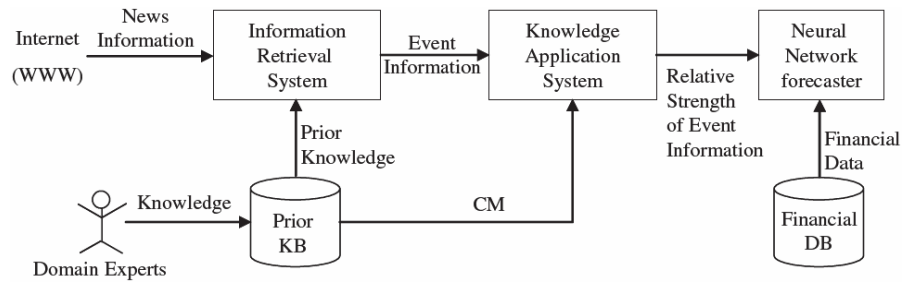
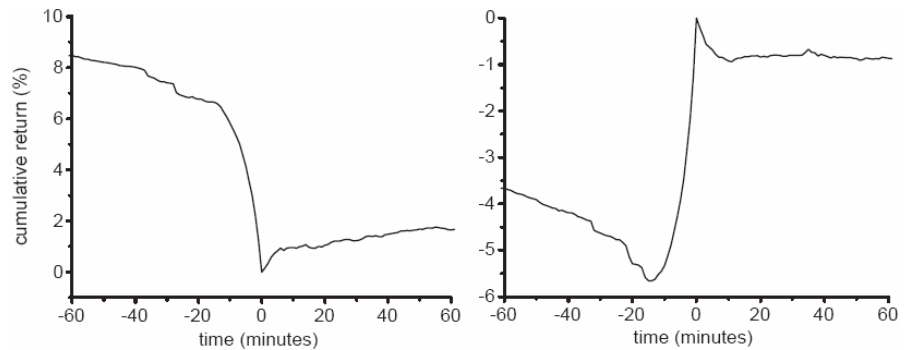


Fig. 3. : Effects of large price changes in short scale [15]



more than 25,000 news and information sources as well as blogs, message boards, and posts from YouTube and Twitter [19].

3.5 RightTrade engine

RightTrade gives investors insight into the media trends. The service provides an interactive and visual representation of real-time media sentiment including the leading global newspapers, business blogs, trade journals, and news wires – across indices, companies, multiple tickers, and customized watch lists. It provides dynamically linked business intelligence information that includes: company profiles, key executives, top competitors, press and public relations activity, quarterly & annual financials, stock quotes, interactive stock charting and SEC filings, news headlines with at-a-glance summaries and sentiment ranking [20].

4 Conclusions

Quantitative interpretation of economic and financial news is studied extensively in the hope of revealing predictive power for financial markets. Impact of news is studied mainly in terms of volatility and trade volume implications. It seems to be empirically true that each category of news has a different effect on volatility and trading activity on different markets. Studies were performed on different time horizons, sometimes resulting in contradictory conclusions.

Experiments show that asset price responses can be ascertained more precisely a short time after the announcement than later in the day. The estimated responses at 4 p.m. exhibit greater standard errors than the responses thirty minutes after the announcement. Moreover, the noise around the predicted responses is also smaller after thirty minutes than later in the day. This result is intuitive: as time passes after a data release, other forces will influence asset prices, increasing the uncertainty of

the response. Asset markets tend to absorb the impact of economic news rather quickly [3].

Adam G. Zawadowski, Gyorgy Andor, and Janos Kertesz analysed the effects of large price changes in short scales in the intraday level. They found overreaction in both cases (sudden price changes and sudden price falls). They concluded that not only the volatility jumps at the event (which itself is the event) but volume, and the bid-ask spread increases as well [15]. The intraday overreaction scenarios suggest that the possible cause of these sudden overreacted intraday events (depicted below) are publication of news. This hypothesis guides me towards further research in this area.

These facts support the view that investors primarily underreact to bad news which suggests an asymmetric response to information [6]. Financial markets become more integrated into world markets, world information becomes relatively more important [4].

Commercialization of research results started and integrated services are evolving that extend financial prediction methods with investor sentiment analysis coming from news sources. Improving the predictive power of investor sentiment by understanding the impact of news by quantitative measures provides prosperous fields for further research.

References

- 1 Tetlock P C, *Giving Content to Investor Sentiment: The Role of Media in the Stock Market*, Journal of Finance **62** (Jun 2007), 1139-1168, DOI 10.1111/j.1540-6261.2007.01232.x;(AN 24998665).
- 2 Tetlock P C, Saar-Tsechansky M, Macskassy S, *More Than Words: Quantifying Language to Measure Firms' Fundamentals*, Journal of Finance **63** (Jun 2008), 1437-1467, DOI 10.1111/j.1540-6261.2008.01362.x;(AN 31961232).
- 3 Bartolini L, Goldberg L, Sacarny A, *How Economic News Moves Markets*, Current Issues in Economics & Finance **14** (Aug 2008).

- 4 **Önder Z, Timga-Muodan Can**, *How Do Political and Economic News Affect Emerging Markets? Evidence from Argentina and Turkey*, *Emerging Markets Finance & Trade* **42** (2006), 50-77, DOI 10.2753/REE1540-496X420403. AN 22173095.
- 5 **Koppel Moshe, Shtrimerberg Itai**, *Good News or Bad News? Let the Market Decide*, Vol. 20, Springer Netherlands, Jan 2006.
- 6 **Chan W S**, *Stock price reaction to news and no-news: drift and reversal after headlines*, *Journal of Financial Economics* **70** (July 2003), 223–260, DOI 10.1016/S0304-405X(03)00146-6.
- 7 **Green T C**, *Economic News and the Impact of Trading on Bond Prices*, *Journal of Finance* **59** (Jun 2004), 1201-1233, DOI 10.1111/j.1540-6261.2004.00660.x; (AN 13017734). 33p, 8 charts, 4 graphs.
- 8 **Jain P C**, *Response of Hourly Stock Prices and Trading Volume to Economic News*, *Journal of Business* **61** (Apr 88), 219-231. 13p; (AN 4584543).
- 9 **Goonatilake R, Herath S**, *The Volatility of the Stock Market and News*, *International Research Journal of Finance and Economics* (2007).
- 10 **Satoru Takahashi, Masakazu Takahashi, Hiroshi Takahashi, Kazuhiko Tsuda**, *Analysis of the Relation Between Stock Price Returns and Headline News Using Text Categorization*, *Lecture Notes in Computer Science*, Vol. 4693, Springer Berlin, Heidelberg, 2007, pp. 1339-1345, DOI 10.1007/978-3-540-74827-4-167, (to appear in print).
- 11 **Liang X**, *Mining associations between web stock news volumes and stock prices*, *International Journal of Systems Science* **37** (2006), no. 13, 919-930, DOI 10.1080/00207720600891562.
- 12 **Balduzzi P, Elton E J, Green T. C**, *Economic News and Bond Prices: Evidence from the U.S. Treasury Market*, *Journal of Financial & Quantitative Analysis* **36** (Dec 2001), 523-543. 21p; (AN 6038946).
- 13 **Taeho Hong**, *Integrated approach of cognitive maps and neural networks using qualitative information on the World Wide Web: the KBNMiner*, *In-goo Han. Expert Systems* **21** (Nov 2004), 243-252, DOI 10.1111/j.1468-0394.2004.00282.x; (AN14927634). 10p, 4 charts, 3 diagrams, 2 bw.
- 14 **Schumaker P R, Chen H**, *Textual Analysis of Stock Market Prediction Using Breaking Financial News: The AZFinText System*, *ACM Transactions on Information Systems* **27** (February 2009), no. 12.
- 15 **Zawadowski Á G, Andor Gy, Kertész J**, *Short-term market reaction after extreme price changes of liquid stocks*, 2004.
- 16 available at http://www.bobsguide.com/guide/prod/5-12205_Reuters_NewsScope_Sentiment_Engine.html.
- 17 available at <http://www.mediasentiment.com/demo/>.
- 18 available at <http://www.yellowbrix.com/index.nsp?sid=bp{\&}pid=3>.
- 19 available at <http://solutions.dowjones.com/economicsentimentindicator/>.
- 20 available at <http://www.righttrade.com/RT/spi/demo>.

Appendix 1

News items were classified into six main groups.

Some real life examples from news and their classification are listed in Table 3.

Appendix 2

Macroeconomic news stories were downloaded from Reuters wire service to track the effects of announcements on asset prices at two time horizons: within thirty minutes of the announcement and at 4 p.m. on the day of the announcement to capture the immediate and full-day response of prices to macroeconomic news.

The nine news releases examined in our analysis encompass thirteen of the nation's most heavily watched economic indicators. The releases (and their corresponding indicators) are as follows:

- 1 Employment Situation Summary (nonfarm payrolls and unemployment rate);
- 2 consumer price index (CPI and CPI excluding food and energy);
- 3 personal income and outlays (personal consumption expenditures [PCE] excluding food and energy, personal income, and personal spending);
- 4 gross domestic product (GDP advance release);
- 5 ISM Manufacturing Report on Business (ISM manufacturing);
- 6 new residential construction (housing starts);
- 7 Conference Board Consumer Confidence Index (consumer confidence);
- 8 University of Michigan Survey of Consumers (consumer sentiment);
- 9 advance monthly sales for retail trade and food services (retail sales less autos).

Impact of announcements of thirteen economic indicators was measured according to their impact on prices in three broad asset classes over a ten-year period ending in 2007:

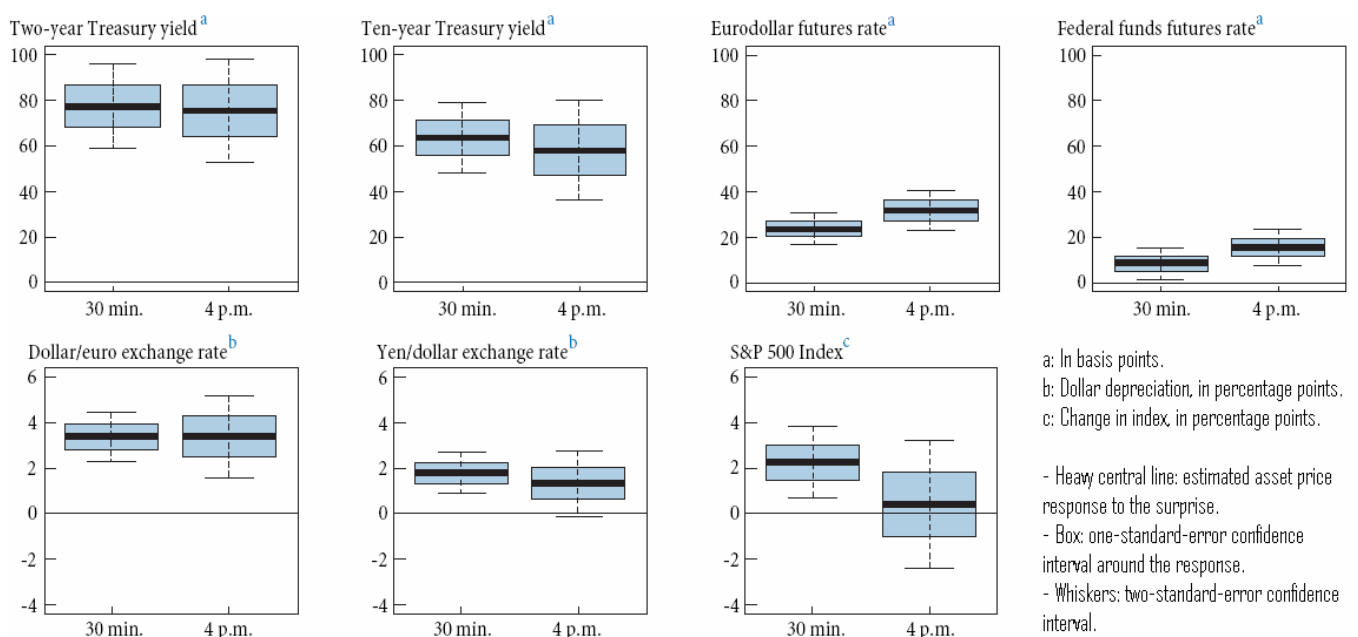
- interest rates;
- foreign exchange rates;
- and equity prices.
- The Employment Situation Summary, with its nonfarm payrolls and unemployment rate components, and the ISM Manufacturing Report on Business have statistically significant effects on five to seven asset prices; the effects are large and persist through the end of the day.

Tab. 2. Classification of news along two dimensions [4]

	World	Domestic	Domestic world
Economic	world economic (affects the whole world)	domestic economic (affects only the particular country)	domestic world economic (affects only the particular country)
Political	world economic (affects the whole world)	domestic economic (affects only the particular country)	domestic world economic (affects only the particular country)

Tab. 3. News examples and their classification where de = domestic economic; dp = domestic political; dwp = domestic world political; dwe = domestic world economic; we = world economic; wp = world political [4]

Date	News type	News
May 28, 1995	de	Argentina privatizes two major railways
June 17, 1995	de	Argentina's tax revenues drop
October 14, 1996	de	Privatization of Etibank
December 30, 1996	de	Turkey's trade deficit widens
May 15, 1995	dp	Menem wins second term as leader in Argentina by a large margin
June 14, 1995	dp	Anti-corruption reporter shot and wounded in Argentina
January 10, 1996	dp	Islamic Party in Turkey is asked to form coalition
December 15, 1996	dp	Release of six Turkish soldiers to Turkey opens door with Kurdish rebels
March 26, 1995	dwp(Argentina)	Argentine arms sold to Ecuador during war with Peru
September 20, 1995	dwp(Argentina)	Britain and Argentina reach an accord of Falkland oil rights
October 8, 1996	dwp (Turkey)	United States criticizes Turkish leader for Libya trip and trade deal
November 28, 1996	dwp (Turkey)	Turkey cancels purchase of ten helicopters from United States
March 19, 1996	dwe (Argentina)	Pride petroleum to buy Argentina drilling contractor
December 23, 1996	dwe (Argentina)	Argentina receives big bank credit line
December 14, 1995	dwe (Turkey)	European parliament admits Turkey to its new customs union
September 5, 1996	dwe (Turkey)	The Overseas Economic Cooperation Fund of Japan extended a loan of 42.31 billion yen (\$ 387.2 million) to Turkey to the second phase of its Istanbul water supply project
June 8, 1995	we	Dollar advances on prospect of interest rate cut by Germany
November 17, 1995	we	600 banks agree to reschedule billions in Russian debt
March 1, 1995	wp	NATO disputes UN reports of possible arms airlift to Bosnia

**Fig. 4.** Impact of macroeconomic news on three different asset classes [3]

Tab. 4. Results for increased volume on bond market after macroeconomic news announcements [7]

	No Announcements 8:00–9:00	Announcements Days		
		Before the Release 8:00–8:30	After the Release	
			8:30–8:45	8:45–9:00
Observations	36563	10182	14739	16120
Std. Dev. of ΔP	1.128	0.957	1.398	1.126
Transactions/hour	65	41	120	131
Volume/transaction	8.0	7.5	9.4	8.9
Volume/hour	515.6	312.0	1126.3	1164.0
Quoted spread	1.690	1.904	1.710	1.552

Tab. 5. Categorization of news stories before measuring the impact on markets [9]

Main Categories	Weighted Factors	subcategories
1. Global stability	4	<ul style="list-style-type: none"> • Terror incidents/reperts/changers/captures • Economic suceses/strenghts • Government new/overthrown Death/assassination of world leaders • Natural man-made disasters
2. Economic data	3	<ul style="list-style-type: none"> • Daily energy prices • Weekly job data interest rates • Lay offs/new hiring • Corporate profits/losses
3. World business	2	<ul style="list-style-type: none"> • Global political successes, announcements, proclamations, or reforms • Corporation mergers/acquisitions • Important international delegations/visits • Major buy-out/financial reports
4. Social-cultural changes	1	<ul style="list-style-type: none"> • Socio-economic status • Major sport championships/victories • Hollywood marriages/divorces • Box-office successes/failures

- The GDP advance release and consumer confidence index also significantly affect most of the asset prices, though for both indicators the number of significant responses drops by day's end.
- Other data releases have a narrower impact. The retail sales indicator, the CPI excluding food and energy, and the consumer sentiment index have notable effects, but the effects are mostly confined to interest rates and diminish perceptibly by day's end.
- The remaining indicators – the CPI, housing starts, and the components of the personal income and outlays report – elicit weak and generally insignificant responses.

As for the distribution of these effects across asset prices, we see that the effects on interest rates and exchange rates tend to persist, while only the unemployment rate has a significant impact on equity prices by day's end, and this impact is not large [3].

Appendix 3

In Green's study announcement surprise is measured as the realized value minus the median survey value. He standardized

surprises to facilitate comparisons across announcements. For announcement type k on day τ , the surprise is defined as follows:

$$S_{k\tau} = \frac{A_{k\tau} - F_{k\tau}}{\sigma_k}$$

where $A_{k\tau}$ and $F_{k\tau}$ are the actual value and median forecast, and σ_k is the standard deviation of $(A_{k\tau} - F_{k\tau})$. Thus, an announcement surprise equal to 1.0 implies a surprise that is one standard deviation greater than zero for that announcement type. The main conclusions of this research are: [7]

- The informational role of trading increases following announcements showing that release of public information increases the level of information asymmetry in government bond markets.
- The price impact varies considerably across announcements with nonfarm payrolls having the largest effect on prices. A one standard deviation positive surprise in Nonfarm Payrolls results in a 28 basis point drop in price. The coefficients for the other announcements are smaller, ranging from 12 basis points for the producer price index to three basis points for initial jobless claims.

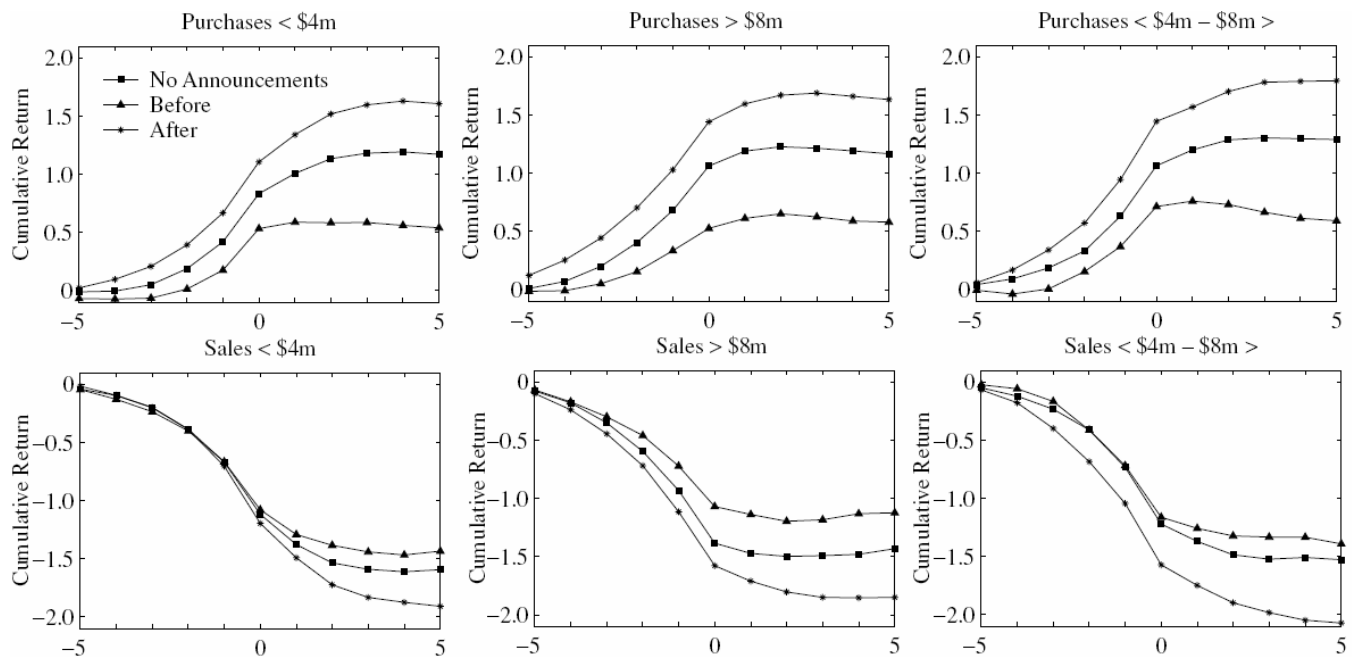


Fig. 5. Cumulated transaction returns around macroeconomic news announcements [7]

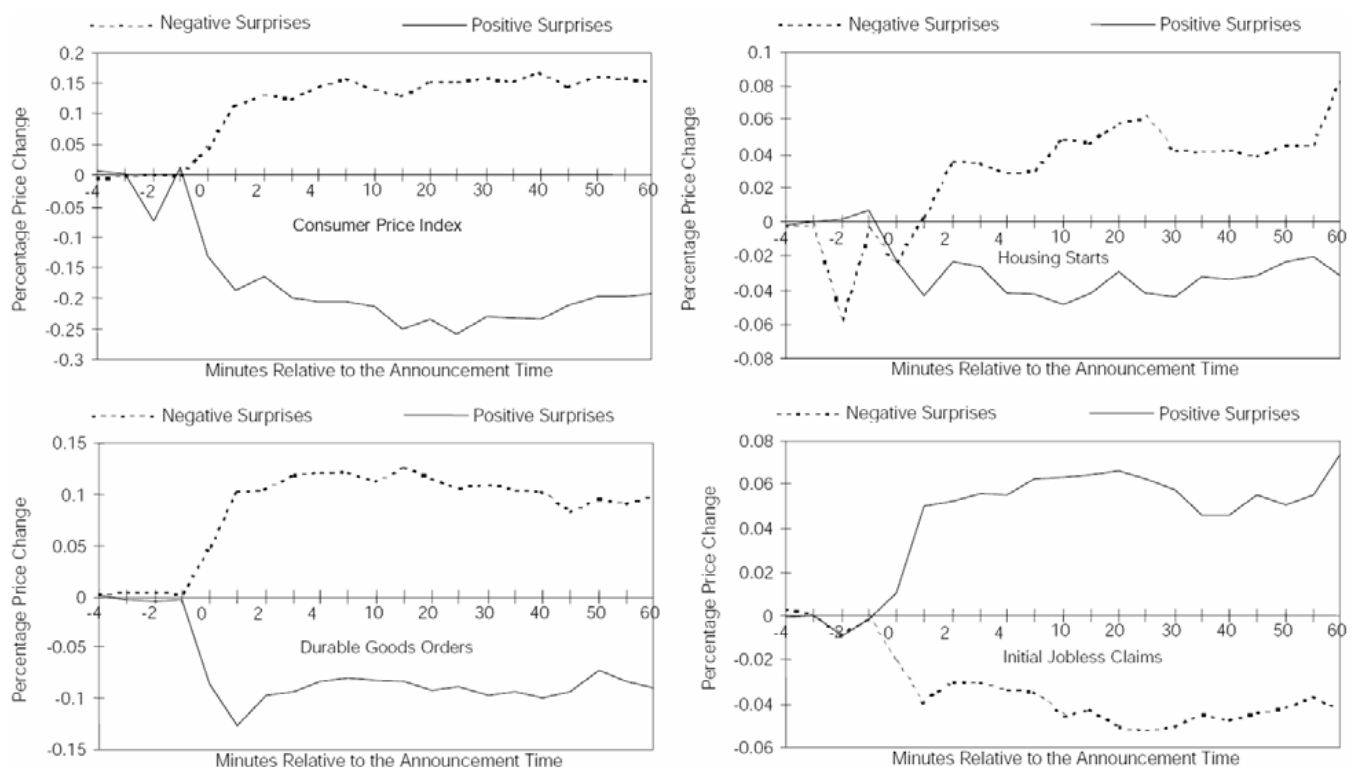


Fig. 6. Intraday bond price reaction to CPI, housing, durable goods orders and initial jobless claim news

- Competition among multiple informed traders leads prices to incorporate private information quickly. This implies that the informational role of trading will decline as time passes after the announcement.

The reason for information asymmetry caused by the phenomena that order flow reveals fundamental information about riskless rates can be originated from three factors:

- Some market participants can have an advantage at determin-

ing how macroeconomic news influences riskless rates.

- Trading volume can increase after a public information event even if market participants interpret the information identically but they had divergent prior expectations.
- The informational role of trading returns to near-normal levels within 15 minutes.

The transaction data was obtained from GovPX that consolidates quote and trade information from the interdealer market

for U.S. Treasury securities and disseminates it through information providers such as Bloomberg. The data on economic announcements is obtained from MMS International (now a subsidiary of Standard and Poor's), a widely used source of forecast data for studies of economic announcements. MMS surveys approximately 40 money market economists each Friday regarding the upcoming week's announcements and records information on the median forecast, the 25th and 75th percentile forecasts, the standard deviation of the forecasts, and the announcement realization.

Results showed that in 15 minutes immediately following the release of news, price volatility increases along with trading intensity and trade size. In the next 15 minutes, trading activity continues to increase while quoted bid-ask spreads narrow. Descriptive data of results can be seen in the Table 4 [7].

Mean cumulative transaction returns were calculated for purchases and sales of different sizes surrounding economic announcements. Trades were sorted into three size groups based on the 33rd and 67th percentiles, which were \$4 million and \$8 million. Cumulative transaction returns can be seen in Fig. 5 measured from five-trades-before to five-trades-after the trade of interest [7].

Appendix 4

The timely reaction for major four macroeconomic news types can be seen in Fig. 6 [12].

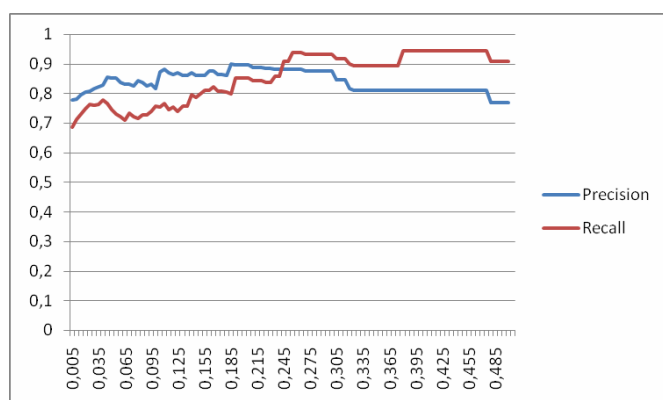


Fig. 7. Precision and recall indicators of my headline news analysis based method

Appendix 5

CNN Headline News Channel, Fox News Channel, www.yahoo.com, and The New York Times (web version-front page) were used as news outlets for the 10-week period of analysis ending on September 10, 2006. News were preclassified items into groups on the basis of their impact on the general public according to Table 5 [9].

Appendix 6

To derive the daily returns from the closing prices I used the following formula, where $R(t)$ is the time dependent return and

$P(t)$ is the time dependent price of the underlying stock.

$$R(T) = \frac{P(T) - P(T - 1)}{P(T - 1)}$$

I created a contingency table from the classification of news and days as follows.

Tab. 6. Categorization of news and days into good and bad categories

	Positive news	Negative news
Positive return	TP (true positive)	FP (false positive)
Negative return	FN (false negative)	TN (true negative)

I used the following indicators to measure the predictive capabilities of my method:

$$Pr = \frac{TP}{TP + FP}; \quad Re = \frac{TP}{TP + FN}$$

where Pr is precision and Re is recall.

I analysed a range of returns (positive and negative) and measured the prediction power of the headline news analysis method. Fig. 7 shows that the method performs at relatively stable 80% precision and recall. The vertical axis shows the precision and recall values in percentage and the horizontal axis shows the values of returns in absolute value (the negative value of return corresponds to days classified as bad and the positive value of the same return corresponds to days classified as good).